

Received August 7, 2019, accepted September 11, 2019, date of publication September 19, 2019, date of current version October 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2942375

Multi-Sensor Depth Fusion Framework for Real-Time 3D Reconstruction

MUHAMMAD KASHIF ALI¹, ASIF RAJPUT^{1,2}, MUHAMMAD SHAHZAD¹, FARHAN KHAN¹, FAHEEM AKHTAR³, AND ANKO BÖRNER²

¹National University of Sciences and Technology (NUST), Islamabad, Pakistan

²German Aerospace Center (DLR), 12489 Berlin, Germany

³College of Software Engineering, Beijing University of Technology, Beijing 100124, China

Corresponding author: Asif Rajput (asif.ali@seecs.edu.pk)

ABSTRACT For autonomous robots, 3D perception of environment is an essential tool, which can be used to achieve better navigation in an obstacle rich environment. This understanding requires a huge amount of computational resources; therefore, the real-time 3D reconstruction of surrounding environment has become a topic of interest for countless researchers in the recent past. Generally, for the outdoor 3D models, stereo cameras and laser depth measuring sensors are employed. The data collected through the laser ranging sensors is relatively accurate but sparse in nature. In this paper, we propose a novel mechanism for the incremental fusion of this sparse data to the dense but limited ranged data provided by the stereo cameras, to produce accurate dense depth maps in real-time over a resource limited mobile computing device. Evaluation of the proposed method shows that it outperforms the state-of-the-art reconstruction frameworks which only utilizes depth information from a single source.

INDEX TERMS 3D reconstruction, LiDAR depth interpolation, multi-sensor depth fusion, stereo vision.

I. INTRODUCTION

Recent advancements in the field of depth sensing systems has made them accessible to the people on a budget, these sensors (Lidar and Stereo) are very significant for 3D Reconstruction. This understanding enables robotic vehicles such as automated drones and underwater vehicles to inspect the areas inaccessible or potentially dangerous for human interaction to this day. To this day, 3D is mostly reconstructed after the vehicle has captured the required environment data but the recent developments in field of computer vision has enabled robotic vehicles with real-time ability to reconstruct a 3D model of the environment with accuracy, hence enabling the field of unmanned area analysis to flourish exponentially.

To reconstruct 3D accurately, a lot of factors such as light conditions, reflective and refractive properties of the surfaces under inspection, weather conditions etc. are to be considered. Usually, an additional noise removal filter is required to cater the additive estimation noise. One can reconstruct 3D using only stereo or LiDAR sensors, but both of these sensors lag in the fields which the other dominates. Stereo sensor systems lag in accuracy of depth sensing but produce

dense measurements with colour information while the Laser based sensor such as LiDAR produce accurate yet sparse depth measurements(as shown in Figure 1), furthermore laser sensing systems are prone to data corruption in multi-sensor environment. This problem motivated various researchers to investigate unique fusion methods to obtain accurate and dense range measurements, hence resulting in various methods containing probabilistic [1], [2] and incremental depth map fusion [3] etc. (discussed in detail in Section III).

In this paper a novel method is proposed to systematically integrate range measurements from stereo and laser based depth sensors followed by reconstructing an accurate 3D model using regularized volumetric integration. The problem of non-uniform 3d samples from LiDAR data has been addressed using a novel multi-stage interpolation method to achieve geometrically accurate dense depth image followed by the depth image fusion with the depth images obtained from stereo by introducing a weighing mechanism designed to handle this particular kind of data and finally produce accurate 3D model.

II. RELATED WORK

State-of-the-art research in Incremental 3D reconstruction techniques which employ depth sensors have reached devel-

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaoqing Pan¹.

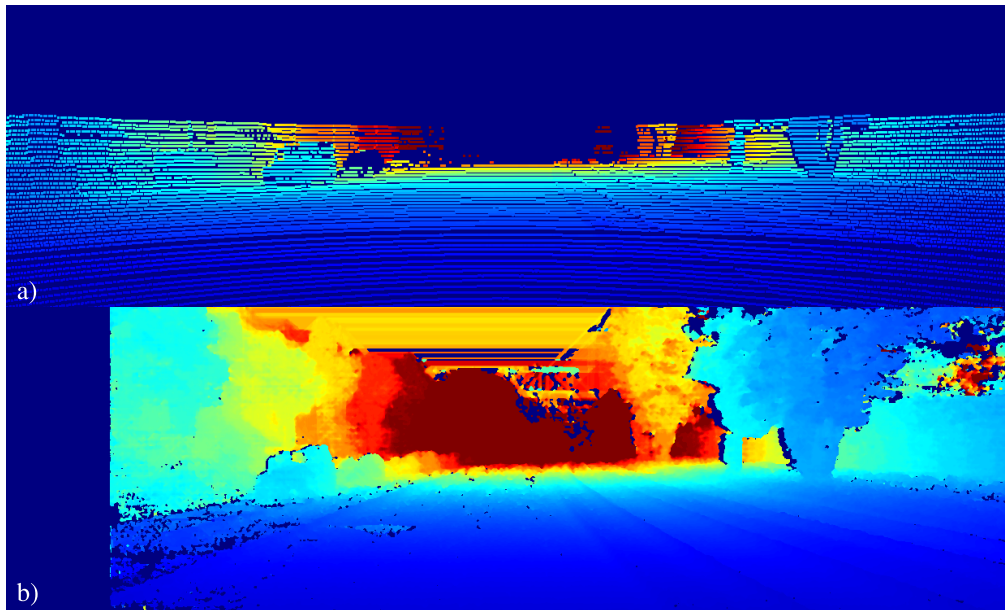


FIGURE 1. a) 3D points from LiDAR depth data projected onto stereo camera and b) depth map estimated from stereo cameras.

opmental plateau. Hackett and Shah [4] demonstrated the benefits of utilizing multi-sensor fusion approach to achieve better geometric understanding of environment. The fusion of LiDAR and Stereo depth is an emerging research domain. Majority of the work in the field of LiDAR and stereo fusion has been done by using a decreased disparity search on stereo data while considering the LiDAR data as ground truth [1], [2]. Huber *et al.* [2] has gone a step further and has combined the reduced disparity search with a dynamic programming framework for faster processing.

A time of flight data and camera fusion has been discussed in [5]. The combination of Airborne LiDAR and satellite imagery has been discussed in [6]–[8] uses LiDAR and aerial imagery fusion to achieve high accuracy surface reconstruction. References [7] and [8] uses this setup to detect and model buildings. This combination of aerial imagery and LiDAR is similar to that of stereo and LiDAR. In both setups, both of the sensors lead in the fields where the other one lags.

The multi-level fusion of LiDAR and Stereo in the field of robotics has been reflected upon in [9] to achieve an obstacle-less path for a mobile robot, while [10] examines a visually accurate 3D reconstruction approach by fusing the colour information obtained through the camera to the depth map obtained by the LiDAR and [2] defines a way of acquiring enhanced disparity images by using a multi-step filter while keeping track of processing time.

Precise temporal association of sensory data from multi-sensor system is an open challenge since problems such as data dropping, sensing latency and bandwidth utilization etc. greatly affect the temporal synchronization of the overall system. Huck *et al.* [11] suggested to use precise timestamping mechanism to tackle unknown delays caused in multi-sensor fusion. Similar approach have been introduced by

Westenberger *et al.* [12] which also incorporates possible drifts of internal sensor clocks as well and determine timestamps up to milliseconds accuracy. Lastly, Kaempchen and Dietmayer [13] suggested to use an intermediate layer for sensor fusion synchronization related issues. Since this paper is focused on 3D reconstruction and aspects related to depth fusion, it is presumed that adequate temporal synchronization strategies have been employed to tackle such temporal association problems in real-life scenarios (readers are encouraged to read official documentation of the KITTI benchmark [19]).

Curless and Levoy [3] introduced a Signed Distance Function (SDF) based volumetric integration method which facilitates representation of sparse depth measurements in a dense bounded voxel space. Rajput *et al.* [14] improved the underlying integration process by introducing a regularized variant least square based fusion which uses a semi-dense voxel space. This efficient utilization of voxel space reduced memory footprint as well as introduced smooth 3D surfaces.

The concept of volumetric fusion and 3D reconstruction is suited specially for relatively small scale environments. State-of-the-art 3D reconstruction techniques (such as [15], [16] and [17]) make use of the volumetric integration process for high quality 3D reconstruction. Unfortunately, the scale of reconstruction based on volumetric integration is bounded with memory constraints and noisy depth data, this problem is addressed by Rajput *et al.* [18] in which a sparse voxel-based approach is employed to extend the 3D reconstruction in boundless fashion.

III. METHODOLOGY

The proposed algorithm (as shown in Figure 2) is designed in a modular structure which allows it to be easily modified or extended as required by the integrated sensor system.

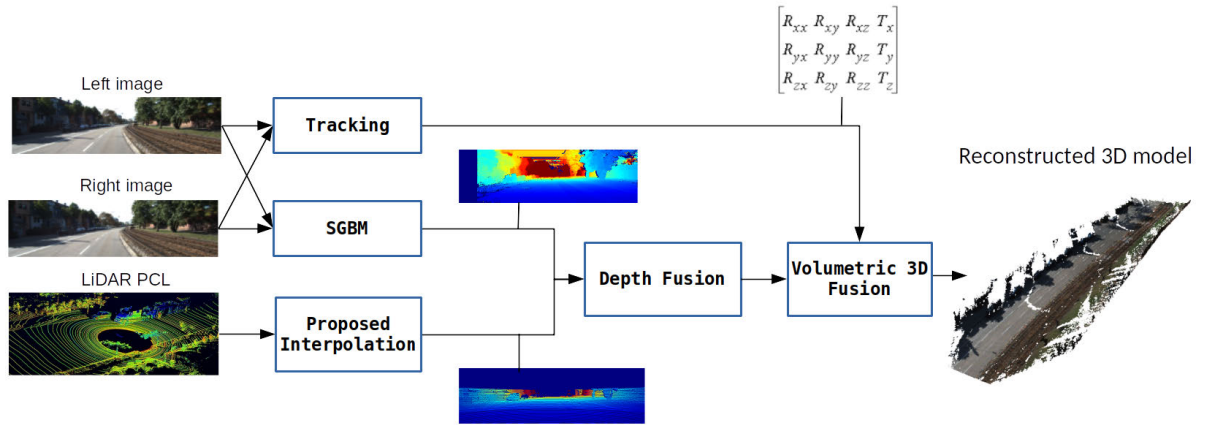


FIGURE 2. Simplified overall structure of the proposed 3D reconstruction framework.

A multi-threaded structure is employed to process each module for faster execution. Tracking, SGBM and the proposed interpolation modules are used to acquire localization, stereo and laser depth data respectively.

Traditional stereo and LiDAR depth fusion approaches represent precise laser data in form of disparity and employ probabilistic methods to fuse data, instead of using disparities, this paper proposes to represent both stereo disparity and laser data in form of depth images. Firstly, the collected data (i.e. Stereo, RGB and LiDAR) is converted to depth data and fed to the Depth Fusion module which melds the input depth maps using the proposed algorithm (as shown in Figure 2) which has been discussed throughout this paper. Volumetric 3D Fusion module assigns the input depth data to spatially accurate three-dimensional point data acquired by the tracking module.

A. CAMERA SETUP

For the purpose of attaining a near accurate 3D model of the environment, AnnyWAY uses multiple actuators for the purpose of acquiring the necessary environment data. These actuators include a rotating laser range sensor for estimating sparse yet accurate depth information, a tetrad of pinhole based HD cameras (two pairs, C_g of grayscale cameras and C_c of conventional RGB cameras for acquiring color definition of the environment) assembled on a railing system to acquire stereo correspondence for disparity estimation, all four of these cameras are separated by a known distance.

The proposed reconstruction framework presumes that the focal length of these cameras is equivalent and known at all times and is denoted by f_c . Since the KITTI Vision Benchmark Suite [19] is a highly documented resource containing calibration information, these intrinsic parameters are used as-is directly from the benchmark suite.

It is also worth mentioning that since extrinsic sensor parameters between LiDAR, C_g and C_c (a.k.a. sensor calibration) were employed to achieve spatial consistency between depth information from LiDAR and stereo camera pair. Furthermore, the system is re-calibrated every time before the

data acquisition as discussed in [19] which ensures that every numerical discrepancy due to wear and tear is captured and the fusion framework uses that information to achieve higher accuracy. Multi-sensor calibration is a crucial research problem which indirectly affects 3D fusion process, however to keep the focus of this research towards 3D reconstruction, readers are encouraged to see [20]–[22]). Similarly, hardware as well as software considerations have been incorporated by Geiger et. al. to tackle synchronization issues between multi-sensor data (readers are encouraged to see Section *Synchronization* from [19]).

At any given time-stamp t , C_g cameras acquire detailed grayscale images Z_t , while the cameras of pair C_c present coloured images I_t , a near accurate but sparse depth map is provided by the Laser range sensor L_t for which a novel interpolation method is proposed which handles the non-uniform 3D samples from LiDAR and produces geometrically accurate depth images, discussed in detail in Section III-B, this information is then fed to “localization” module which estimates sensor movements and generates sensor pose in world coordinate system consisting of translation $T_t \in \mathbb{R}^{3 \times 3}$ and orientation $R_t \in SO(3)$.

Therefore, A pixel $[x, y]^T$, on reconstructed model can be related to a global 3D point $P_w \in \mathbb{R}^{3 \times 3}$ by

$$P_w = R_t \cdot \begin{bmatrix} (row - c_x) \frac{Z_t(row, col)}{f_x} \\ (col - c_y) \frac{Z_t(row, col)}{f_y} \\ Z_t(row, col) \end{bmatrix} + T_t \quad (1)$$

Additional scaling can be applied to achieve multi-scaled reconstruction. However, in this case where the environment is relatively large, a fixed scale is selected at the time of execution. The images provided by C_g are used to produce a disparity map for depth estimation which is then merged with the depth perceived by LiDAR. Since the main goal of this paper is to achieve an efficient 3D model of the surroundings, a novel Fusion mechanism is contemplated to efficiently merge the depth representations obtained from

both the cameras and LiDAR which is discussed in detail in Section III-C.

B. COMPOUND INTERPOLATION

The depth perception provided by the LiDAR is in the form of a point cloud and it contains some gaps due to its rotating effect of the LiDAR, these gaps in the point cloud can be approximated by using curve fitting techniques. In this paper, different interpolation methods are compared and a novel interpolation method is presented, which combines the strengths of the best available approximation techniques in the present time.

Firstly, Linear Interpolation (LI) was considered to acquire a geometrically accurate 3D model from the point cloud, due to LI's unswerving nature it produced some steep sloped edges but despite these edges, LI yielded sound results across relatively longer gaps.

$$P(x) = P_0 + \frac{(x - x_0)(P_n - P_0)}{(x_n - x_0)} \quad (2)$$

Secondly, the Cubic Spline Interpolation (CSI) method was considered, due to CSI's cubic factors, it rendered extreme curvatures across longer frames but worked near perfectly in relatively shorter intervals. The CSI contains more arithmetic operations than the LI, therefore it is a sluggish process.

$$\begin{aligned} P(x) &= \sum_{i=0}^3 a_i(x - x_0)^i \\ &= a_0(x - x_0)^0 + a_1(x - x_0)^1 \\ &\quad + a_2(x - x_0)^2 + a_3(x - x_0)^3 \end{aligned} \quad (3)$$

In the above equation, a_0 through a_3 are the co-efficients, evaluated by using the basic interpolation properties which state that near the start and end point of the frame of consideration, the slope should be negligible and at the extreme points of the frame, the function is equal to the values of extremes to ensure smoother curves.

$$\begin{aligned} P(x_0) &= P(x_n) = 0 \\ P(x_0) &= P_0 \\ P(x_n) &= P_n \end{aligned} \quad (4)$$

Through mathematical manipulation of the basic equation using the above properties, a system of equations is acquired, containing 4 equations and unknowns (a_0 , a_1 , a_2 , and a_3). By solving the system of equations, the unknowns are evaluated as:

$$\begin{aligned} a_0 &= P_0 \\ a_1 &= 0 \\ a_2 &= \frac{\alpha(P_1 - P_0)}{\beta(x_n - x_0)^2} \\ a_3 &= \frac{(P_0 - P_1)}{(x_n - x_0)^3} \end{aligned} \quad (5)$$

where values of α and β are used to modify the weights assigned to control the parameters of approximation. It was

found with extensive empirical evaluation that the system performed relatively better with $\alpha = 2$ and $\beta = 3$. After carefully analyzing the characteristics of both the interpolation techniques discussed, a novel approximation method is proposed, which combines the strength of both CSI and LI. The proposed method processes the point cloud in horizontal and vertical iterations. Firstly, a frame length in a straight line is construed. If the length defined is relatively small, CSI is used and for the longer lengths LI is used. A point cloud when sliced in a straight line, can be considered as a two-dimensional graph with the incrementing index along that line as a baseline axis and the point depths on that baseline axis can be considered as the vertical axis values. By using this mechanism, a point cloud can be sliced both horizontally and vertically for easier and faster calculations.

Since each parallel plane slice is independent, a multi-threaded processing platform can be used to make the calculations more expeditious. A single iteration of the Gaussian filter discussed, is adopted to smoothen the recreated surface and to fill up small gaps remained in the point cloud under consideration. This filter, when used with smaller gaps between the pixels, approximates the unknown pixels pretty accurately.

An iterative Gaussian filter based approximation technique was also examined, it assigned the average value of the immediate nearby pixels to the middle pixel. A point cloud acquired through a LiDAR usually contains bigger gaps than one pixel, so the reconstructed surface turns out to be smudgy with uneven patterns.

C. REGULARIZED VOLUMETRIC 3D FUSION

Traditional volumetric integration proposed in [3] is capable of integrating multiple depth samples to facilitate multi-view and temporal updates to the overall reconstruction. Underlying integration principle is fairly simple in which a volumetric grid G is subdivided into a uniform bounding boxes (commonly referred to as voxels). Each interest voxel $v \in G$ is processed with a function $f(v) : \mathbb{R}^3 \rightarrow \mathbb{R}^1$ which transforms spatial information of the voxel into an expected signed distance function followed by weighted integration which accommodates incremental updates to overall reconstruction. Traditional 3D reconstruction frameworks employ weighted integration of each incremental update to exploit stochastic convergence property, however this exploitation depends greatly on the number of updates and Rajput *et al.* [23] showed that sensors (such as passive depth and LiDAR sensors) with lower sensing frequency are prone to produce noisy surfaces.

This problem of reducing depth noise at the time of integration is addressed in [14] where a total variation filtering based regularization is employed to reduce the effects of noise in recursive manner. It is possible to treat both depth images (i.e. estimated from stereo camera and interpolated from proposed pipeline) as temporal updates and fuse them sequentially, however such integration requires multiple updates to volumetric grid which results in poor computational efficiency.

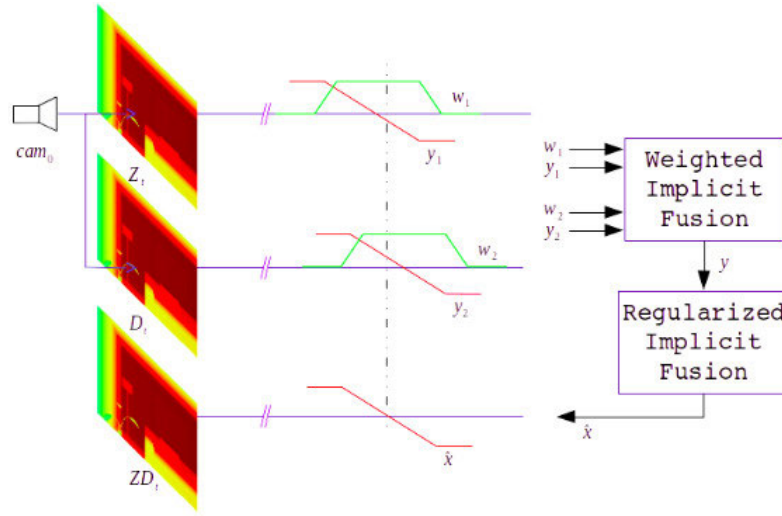


FIGURE 3. Illustration of the proposed 3D volumetric fusion process.

In contrast, we propose a novel two stage integration system which fuses two depth images in the first stage and employs total variation filtering based regularization to perform the implicit smoothing(as shown in Figure 3).

Initially, both depth images are traversed in raster scanning method simultaneously and each depth sample are represented as a pair of SDF-signal and weight signal (i.e. $s_1 = \{y_1, w_1\}$ and $s_2 = \{y_2, w_2\}$) which are processed with following integration function

$$y = \frac{(y_1 w_1) + (y_2 w_2 w_1)}{w_1 + (w_2 w_1)} \quad (6)$$

It is worth mentioning that the process of multiplying \hat{w}_1 with contents of s_2 enforces an additional constraint which validates the contents of s_2 to be in specified range. In a special case where depth sample $z(\text{row}, \text{col})$ is invalid, the weighted integration system is designed to use $w_1 = 1$ to ensure that depth value from stereo depth image is utilized. Secondly, the calculated signal y is processed with regularized implicit fusion module which treats y as a noisy measurement for least squares system while the estimated state of system (i.e. \hat{x}) is expected to exhibit smoother implicit iso-surface. Such minimization system can be written in following minimization system

$$\hat{x} = \arg \min_x \{\|x - y\|^2 + \lambda \|g(x)\|^2\} \quad (7)$$

where λ is the regularization parameter which controls the influence of neighbouring elements for SDF-signal. The mathematical solution to Equation 6 is used as-is from [14] where a recursive variant of the least squares system is derived and implemented. Finally, the estimated values of x are updated within the global voxel-grid.

D. VALUE FOR THE REGULARIZATION PARAMETER

As discussed in the Section III-C that the concept of regularization is integrated within the fusion framework to

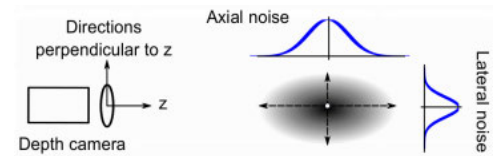


FIGURE 4. A 3D noise distribution of Kinect depth measurement in terms of axial (z-direction) and lateral (directions perpendicular to z) noise. [25].

reduce the effects of noise while producing smooth surfaces. Rajput *et al.* [24] argued that addition of smoothing constraint within 3D fusion frameworks also accelerates the incremental integration process. Therefore, selecting appropriate value(s) for the regularization parameter λ is essential and plays a vital role in the overall reconstruction process.

It is a well established phenomenon that the process of sensing involves the addition of some additive noise to the actual value(s). In the case of depth sensing using Microsoft Kinect camera, Nguyen *et al.* [25] suggested to categorize depth noise in terms of lateral and axial noise as shown in Figure 4. They discovered that both the lateral and axial noise can be approximated using Gaussian distributions. In principal, acquiring descriptive parameters (such as mean μ and standard deviation σ) for both distributions can be determined by using a 3D model of pre-defined environment (i.e. ground-truth) and abundant depth observations from various locations. Similar findings have been reported in detail by Choo *et al.* [26], however detailed description on such findings is out of the scope for this paper. Unfortunately, setting up such elaborate set-up for sensor noise parameters require both tedious empirical evaluation and does not generalize well for different sensor. Instead, generalized properties of the regularization parameter can be more useful than specialized noise profiling for every depth sensor.

In order to acquire such generalized trends of smoothing process at the time of incremental integration process.

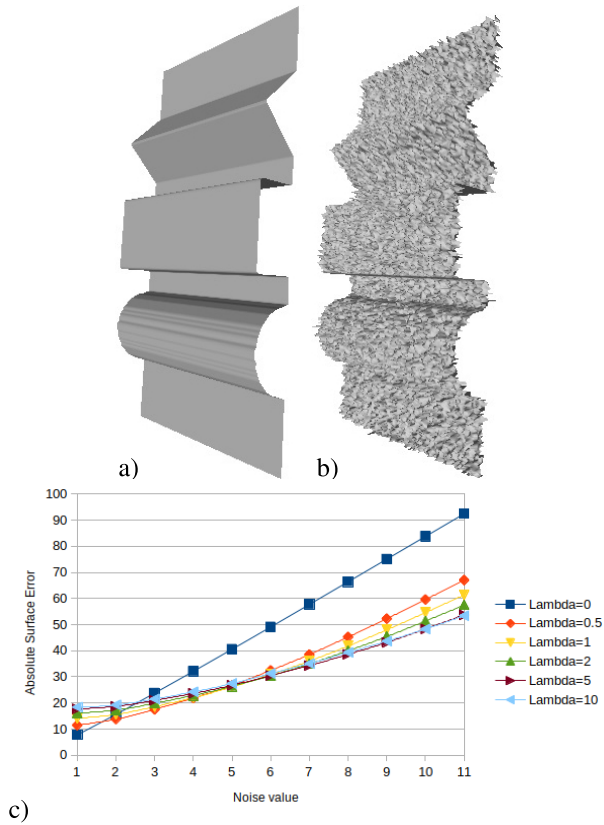


FIGURE 5. a) Synthetic 3D surface, b) Synthetic surface corrupted with noise and c) Effects of λ with incremental depth noise in millimeters (lower is better).

TABLE 1. Disparity error calculations.

	Foreground	Background	All
Stereo (SGBM) [27]	12.62%	18.93%	13.58%
Fused stereo	4.20%	9.15%	4.94%

A synthetic 3D surface consisting of smooth, edges and sharp boundaries (as shown in Figure 5.a) is projected onto a virtual camera and respective precise depth is recorded. Acquired depth values (in the form of a depth image) are then corrupted with various degrees of noise and fused together using different values of λ to emulate 3D fusion process. Absolute error of the incremental fusion are accumulated and a detailed empirical results were acquired and are shown in Figure 5.c.

It is evident from the provided analysis in Figure 5.c that when the added noise is smaller, lower λ values perform well relatively compared to λ values. This observation is aligned with the fact that high quality depth information does not require any external smoothing, in-fact applying smoothing to such high quality information will potentially degrade the reconstruction at sharp edges etc. Similarly, a correlated relation of noise with smoothing effect can be observed when the added noise is relatively higher. In such scenarios it can be observed that there exists a correlation between higher values of noise with higher λ values.

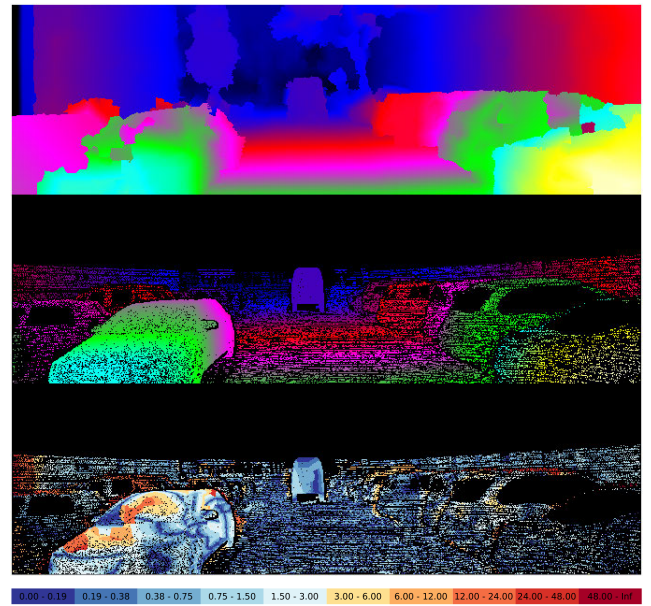


FIGURE 6. Estimated disparity image $d_{est}(x, y)$, ground truth disparity $d_{gt}(x, y)$ and error image (top, middle and bottom row respectively).

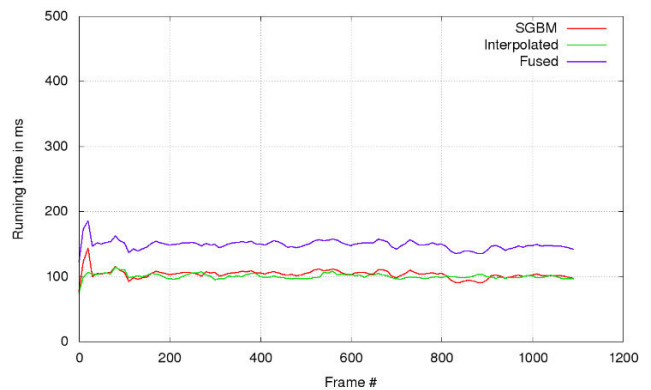


FIGURE 7. Running time analysis for reconstructing KITTI sequence 06.

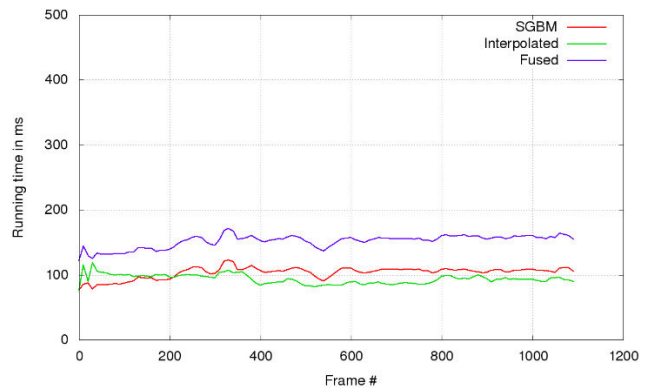


FIGURE 8. Running time analysis for reconstructing KITTI sequence 07.

In conclusion, an optimal solution to finding appropriate λ values for each acquired depth requires sensor noise profiling and tedious set-up which is only feasible for specific scenarios. Therefore, it is strongly suggested to perform noise

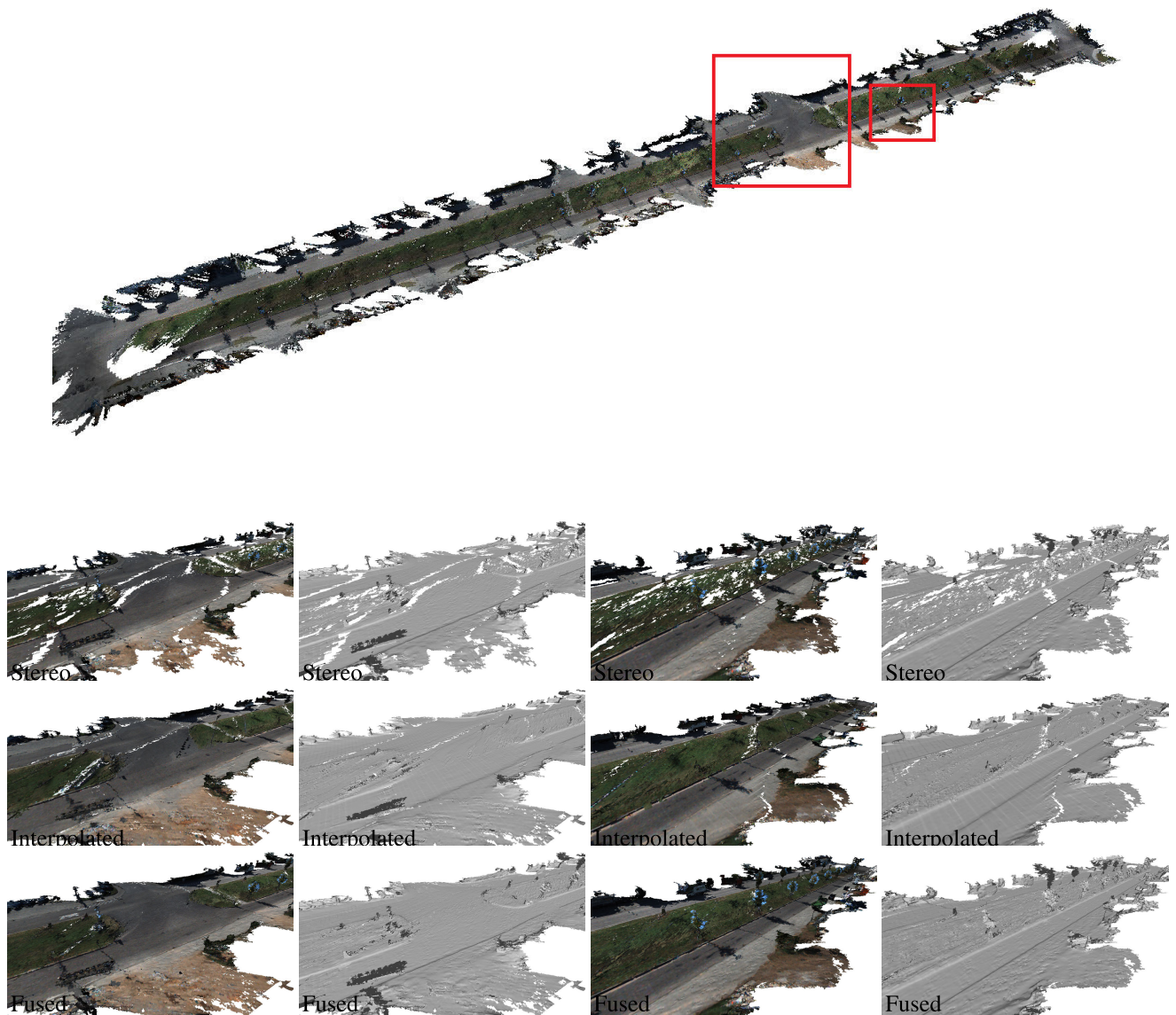


FIGURE 9. Reconstructed model from KITTI dataset, sequence 06.

profiling beforehand, however in the case of *KITTI* dataset we have used $\lambda = 5$ which produced overall lower absolute surface error.

IV. EXPERIMENTAL SETUP

A. HARDWARE

The datasets used for the experimental evaluation of the presented algorithm has been acquired by “Kitti” with HDL-64E laser sensor which captures 100k points per frame and ten frames per second with a vertical resolution of 64. The stereo cameras used are also triggered at the rate of 10 frames per second with dynamic shutter adjustment and the resolu-

tion of the image received through each camera is of 1382×512 pixels. The baseline distance of these cameras is around 55 cm and the focal length is about 750 pixels.

B. SOFTWARE

The creation of detailed meshes require a huge amount of input data containing point clouds, stereo images for depth and color definition, location and orientation data. The *KITTI* dataset contains 15 sequences each containing approximately 1100 captured data instances (accumulating up to of 167 gigabyte of data containing four high-quality images, LiDAR point cloud, accelerometer and electronic compass data for

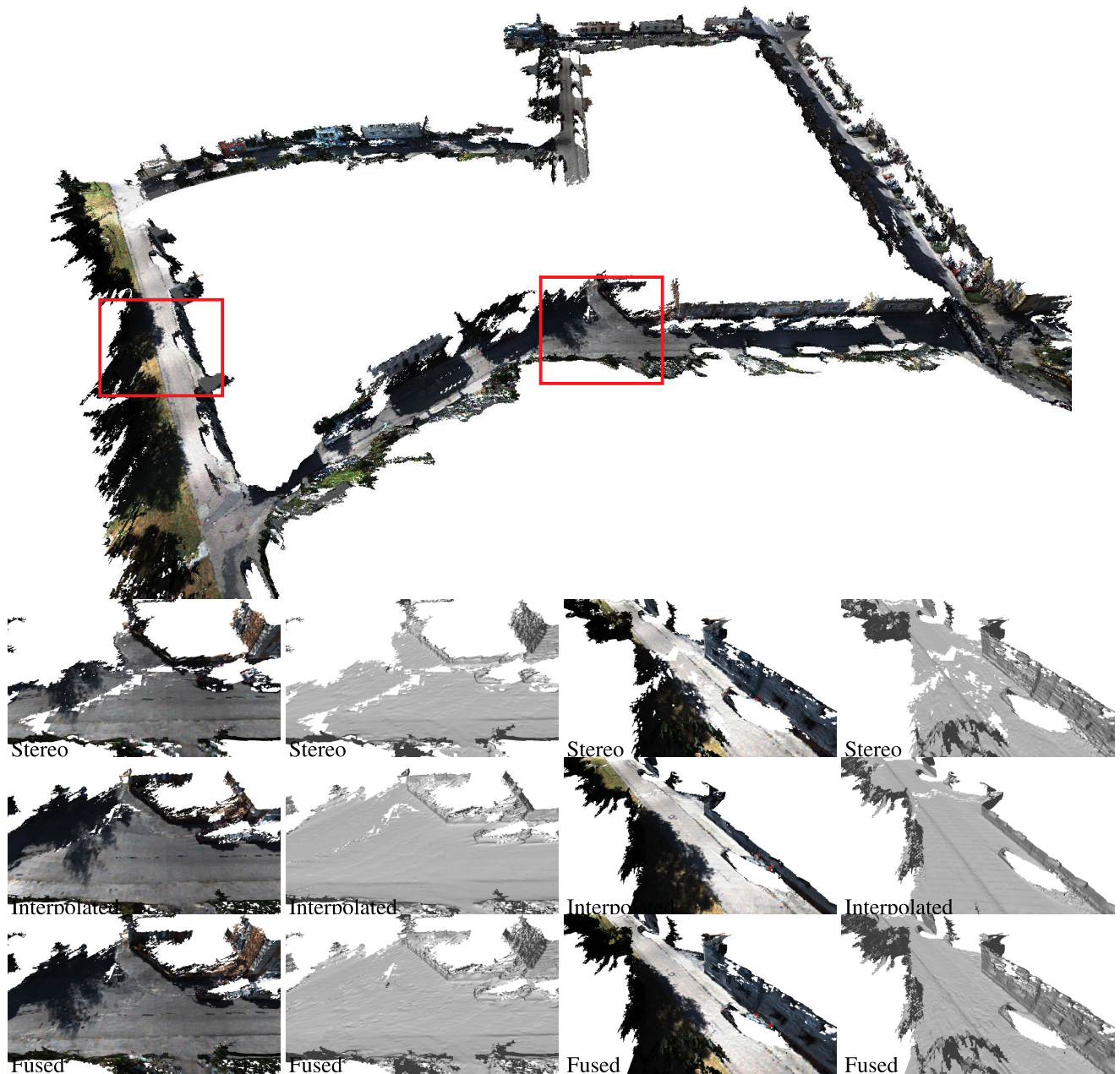


FIGURE 10. Reconstructed model from KITTI dataset, sequence 07.

each data instance). In order to summarize our findings without presenting repetitive qualitative results, sequence 06 and 07 were selected since they contain relatively easy and hard environments respectively. Clearly processing time and managing space are critical factors for the execution of the proposed algorithm in real-time.

C. EVALUATION

Proposed method is tested thoroughly and comparative results are presented in this section to highlight quantitative and qualitative results. Following methods are employed on both of the dataset trajectories (Kitti sequence 06 and 07):

- SGBM.
- Interpolated.
- Fused (Proposed).

In a traditional evaluation scenario in which ground-truth 3D model is available for inspection, quantitative metrics such as absolute surface error, mean, standard deviation and median etc can be calculated. However, realistic nature of the acquired data combined with large scale environment characteristics restricts such quantitative evaluation. Fortunately, KITTI dataset comes with a development kit to evaluate disparity errors from stereo images using synthesized *ground-truth* disparity images from LiDAR data. The development kit

measures disparity errors using:

$$\varepsilon = \frac{|d_{gt}(x, y) - d_{est}(x, y)|}{N} \quad (8)$$

where $d_{gt}(x, y)$ and $d_{est}(x, y)$ are ground truth disparity images and estimated disparity respectively and N is number of valid disparity within $d_{gt}(x, y)$. Figure 6 highlights the underlying process of error estimation in a pictorial format. Aforesaid disparity error is further divided into *background*, *foreground* and *all* to highlight properties of disparity estimation in a more comprehensive metric. All relevant disparity images have been processed with the development kit and results of these quantitative evaluation are presented in Table 1 where it can be seen that the proposed fusion approach reduced the disparity error in both foreground and background disparities. It is worth mentioning that the provided quantitative results of SGBM [27] are ranked at 208th position while the ranking is improved to 159th. Since the evaluation benchmark does not support or provide results for multi-sensor disparity errors, it was decided to deliberately avoid comparing the proposed technique with pure stereo matching techniques and/or networks. Similar improvements are also expected to exhibit while applying to the state-of-the-art stereo matching approaches.

Screenshots of reconstructed 3D models of trajectories (please see Figure 9 and 10) are provided to facilitate visual inspection and analysis to evaluate performance overall system in a qualitative manner. It can be observed from Figures 9 and 10 that fusion of interpolated LiDAR data with depth images from stereo camera system produced high quality meshes. It is worth mentioning that 3D meshes generated from stereo depth image suffer greatly in texture-less surfaces (such as roads, walls etc). Fortunately, due to hybrid nature of multi-sensor system, interpolated LiDAR depth images are unaffected of these problems and hence the resulting fused 3D meshes contain greater surface area compared to using either stereo or LiDAR data (these effects can be seen in the zoomed in sub-figures of Figures 9 and 10).

All experimentation is carried on machine having following specifications:

- Intel Core i7-4790.
- Nvidia Quadro K620.¹
- 8GB RAM.
- Windows 7 (64-bit) and Linux 14.04 Operating System.

Figure 7 and 8 represent the execution time taken by all the methods to integrate sequences 06 and 07 respectively.

According to the provided running time analysis, it was concluded that the proposed method after integrating uses around 50ms more than the other two and provides geometrically accurate meshes and the execution time taken is almost unaffected by the size of the dataset. This processing time can be further reduced by using a multithreaded mechanism for running the algorithm according to the interval in which a new batch of data is acquired.

¹Used only for SGBM

V. CONCLUSION

In this paper, we presented a novel method to produce high quality 3D models of small and large scale environments. This method utilizes two different depth measuring sensors, LiDAR and stereo cameras to produce the 3D models. The problem of sparse point clouds acquired through the LiDAR has been tackled by using a combination of various state of the art approximation techniques and the processed, dense geometrically accurate point clouds are then merged with the depth maps obtained from the stereo cameras.

This merger is achieved by implementing a unique weighing mechanism in which the net depth is calculated through assigning dynamic weights to both depth maps, under consideration. Fused method outperformed the commonly used, high-end methods for 3D reconstruction while being computationally inexpensive. By using the multi-threaded architecture of the modern-day CPU, this method can be used to produce 3D meshes in real-time.

ACKNOWLEDGMENT

Authors would like to appreciate insights and suggestions of anonymous reviewers which enhanced quality as well as readability of the presented research.

REFERENCES

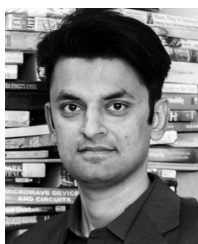
- [1] W. Maddern and P. Newman, "Real-time probabilistic fusion of sparse 3D LiDAR and dense stereo," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 2181–2188.
- [2] H. Badino, D. Huber, and T. Kanade, "Integrating LiDAR into stereo for fast and improved disparity computation," in *Proc. Int. Conf. 3D Imag., Modeling, Process., Vis. Transmiss.*, May 2011, pp. 405–412.
- [3] B. Curless and M. Levoy, "A volumetric method for building complex models from range images," in *Proc. 23rd Annu. Conf. Comput. Graph. Interact. Techn.*, 1996, pp. 303–312.
- [4] J. K. Hackett and M. Shah, "Multi-sensor fusion: A perspective," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 1990, pp. 1324–1330.
- [5] V. Gandhi, J. Čech, and R. Horaud, "High-resolution depth maps based on TOF-stereo fusion," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2012, pp. 4742–4749.
- [6] T. Schenk and B. Csathó, "Fusion of LiDAR data and aerial imagery for a more complete surface description," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 34, no. 3/A, pp. 310–317, 2002.
- [7] L. C. Chen, T.-A. Teo, Y.-C. Shao, Y.-C. Lai, and J.-Y. Rau, "Fusion of LiDAR data and optical imagery for building modeling," *Int. Arch. Photogramm. Remote Sens.*, vol. 35, no. B4, pp. 732–737, 2004.
- [8] G. Sohn and I. Dowman, "Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction," *ISPRS J. Photogramm. Remote Sens.*, vol. 62, no. 1, pp. 43–63, 2007.
- [9] K. Nickels, A. Castano, and C. M. Cianci, "Fusion of LiDAR and stereo range for mobile robots," in *Proc. Int. Conf. Adv. Robot. (ICAR)*, Jun./Jul. 2003, pp. 65–70.
- [10] A. Harrison and P. Newman, "Image and sparse laser fusion for dense scene reconstruction," in *Field and Service Robotics*. Berlin, Germany: Springer, 2010, pp. 219–228.
- [11] T. Huck, A. Westerberger, M. Fritzsche, T. Schwarz, and K. Dietmayer, "Precise timestamping and temporal synchronization in multi-sensor fusion," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 242–247.
- [12] A. Westerberger, T. Huck, M. Fritzsche, T. Schwarz, and K. Dietmayer, "Temporal synchronization in multi-sensor fusion for future driver assistance systems," in *Proc. IEEE Int. Symp. Precis. Clock Synchronization Meas., Control Commun.*, Sep. 2011, pp. 93–98.
- [13] N. Kaempchen and K. Dietmayer, "Data synchronization strategies for multi-sensor fusion," in *Proc. IEEE Conf. Intell. Transp. Syst.*, vol. 85, Nov. 2003, pp. 1–9.

- [14] M. A. A. Rajput, E. Funk, A. Börner, and O. Hellwich, "3D virtual environments and surveillance applications ; image and video processing, compression and segmentation," in *Proc. 13th Int. Joint Conf. e-Bus. Telecommun.*, vol. 5, 2016, pp. 72–80. doi: [10.5220/0005967700720080](https://doi.org/10.5220/0005967700720080).
- [15] F. Steinbrücker, J. Sturm, and D. Cremers, "Volumetric 3D mapping in real-time on a CPU," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May/Jun. 2014, pp. 2021–2028.
- [16] O. Kähler, V. A. Prisacariu, C. Y. Ren, X. Sun, P. Torr, and D. Murray, "Very high frame rate volumetric integration of depth images on mobile devices," *IEEE Trans. Vis. Comput. Graphics*, vol. 21, no. 11, pp. 1241–1250, Nov. 2015.
- [17] T. Whelan, S. Leutenegger, R. F. Salas-Moreno, B. Glocker, and A. J. Davison, "ElasticFusion: Dense SLAM without a pose graph," in *Proc. 11th Robot., Sci. Syst. (RSS)*, Jul. 2015, pp. 1–9. doi: [10.15607/RSS.2015.XI.001](https://doi.org/10.15607/RSS.2015.XI.001).
- [18] M. A. A. Rajput, E. Funk, A. Börner, and O. Hellwich, "Boundless reconstruction using regularized 3D fusion," in *Proc. Int. Conf. E-Bus. Telecommun.* Cham, Switzerland: Springer, 2016, pp. 359–378.
- [19] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, Sep. 2013.
- [20] D. Zuñiga-Noël, J.-R. Ruiz-Sarmiento, R. Gomez-Ojeda, and J. Gonzalez-Jimenez, "Automatic multi-sensor extrinsic calibration for mobile robots," *IEEE Robot. Autom. Lett.*, vol. 4, no. 3, pp. 2862–2869, Jul. 2019.
- [21] J.-C. Devaux, H. Hadj-Abdelkader, and E. Colle, "A multi-sensor calibration toolbox for Kinect: Application to Kinect and laser range finder fusion," in *Proc. 16th Int. Conf. Adv. Robot. (ICAR)*, Nov. 2013, pp. 1–7.
- [22] R. Unnikrishnan and M. Hebert, "Fast extrinsic calibration of a laser rangefinder to a camera," *Robot. Inst.*, Pittsburgh, PA, USA, Tech. Rep. CMU-RI-TR-05-09, 2005.
- [23] A. Rajput, E. Funk, A. Börner, and O. Hellwich, "A regularized volumetric fusion framework for large-scale 3D reconstruction," *ISPRS J. Photogramm. Remote Sens.*, vol. 141, pp. 124–136, Jul. 2018.
- [24] M. A. A. Rajput, "A regularized fusion based 3D reconstruction framework: Analyses, methods and applications," Tech. Univ. Berlin, Berlin, Germany, Tech. Rep., 2018. doi: [10.14279/depositonce-7471](https://doi.org/10.14279/depositonce-7471).
- [25] C. V. Nguyen, S. Izadi, and D. Lovell, "Modeling kinect sensor noise for improved 3D reconstruction and tracking," in *Proc. 2nd Int. Conf. 3D Imag., Modeling, Process., Vis. Transmiss.*, Oct. 2012, pp. 524–530.
- [26] B. Choo, M. Landau, M. DeVore, and P. A. Beling, "Statistical analysis-based error models for the microsoft Kinect depth sensor," *Sensors*, vol. 14, no. 9, pp. 17430–17450, 2014.
- [27] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2007.



MUHAMMAD KASHIF ALI received the B.E. degree in electrical engineering from the Pakistan Navy Engineering College (PNEC), Karachi, in 2018, which is a constituent college of the National University of Sciences and Technology (NUST), Islamabad.

His research interests include image processing, computer vision, 3D reconstruction, and artificial intelligence.



ASIF RAJPUT received the B.E. degree in computer system engineering from QUEST Nawabshah, in 2009, the M.Sc. degree in computer engineering from the College of Electrical and Mechanical Engineering (CEME), a constituent college of the National University of Sciences and Technology (NUST), in 2011, and the Ph.D. degree by research from the Technical University of Berlin, where he conducted his research work with German Aerospace Center (DLR).

His research interests include digital image processing, computer vision, machine learning, and deep neural networks.



MUHAMMAD SHAHZAD received the B.E. degree in electrical engineering from the National University of Sciences and Technology, Islamabad, Pakistan, in 2004, the M.S. degree in autonomous systems from the Bonn-Rhein-Sieg University of Applied Sciences, Sankt Augustin, Germany, in 2011, and the Ph.D. degree in radar remote sensing and image analysis from the Department of Signal Processing in Earth Observation (SiPEO), Technische Universität München (TUM), Munich, Germany, in 2016. He attended two two weeks professional thermography training course at the Infrared Training Center, North Billerica, MA, USA, from 2005 to 2007.

He was a Visiting Research Scholar with the Institute for Computer Graphics and Vision, Technical University of Graz, Austria. Since 2016, he has been a Senior Researcher with SiPEO, TUM, Germany, and an Assistant Professor with the School of Electrical Engineering and Computer Science, National University of Sciences and Technology. He has been with the Department of Electronic and Power Engineering, NUST-PNEC, since 1998, where he is currently an Associate Professor. He is also a Co-Principal Investigator of the recently established Deep Learning Laboratory (DLL) under the umbrella of National Center of Artificial Intelligence (NCAI), Islamabad. His research interests include deep learning for processing unstructured/structured 3-D point clouds, optical RGBD data, very high-resolution radar images, reconfigurable computing, cryptography, FPGA-based system design, information security, cryptographic engineering, and software defined networking.



FARHAN KHAN received the B.S. degree (Hons.) in electrical and electronics engineering from the University of Engineering and Technology, Peshawar, Pakistan, in 2007, the M.Sc. degree in RF communication systems from the University of Southampton, U.K., in 2009, and the Ph.D. degree in electrical and electronics engineering from Bilkent University, Ankara, Turkey, in 2017. He was a Lecturer with the Department of Electrical Engineering, COMSATS Institute of Information Technology, Pakistan, from 2007 to 2013. He was a Postdoctoral Fellow with the Singapore University of Technology and Design, as a Data Scientist. He is currently an Assistant Professor with the School of Electrical Engineering and Computer Science (SEECs), National University of Sciences and Technology Islamabad, Pakistan. His research interests include adaptive signal processing, data science, machine learning, industrial informatics, and neural networks.



FAHEEM AKHTAR received the M.S. degree in computer science from the National University of Computing and Emerging Sciences (Fast NUCES), Karachi, Pakistan. He is currently pursuing the Ph.D. degree with the School of Software Engineering, Beijing University of Technology, Beijing, China. He is also an Assistant Professor with the Department of Computer Science, Sukkur IBA University, where he is on study leave. He is author of various SCI, EI, and Scopus indexed journals and international conferences. His research interests include data mining, machine learning, information retrieval, privacy protection, internet security, the Internet of Things, and big data.



ANKO BÖRNER received the degree in electrical engineering from the University of Ilmenau, and the Ph.D. degree from the German Aerospace Center (DLR).

He was a Postdoctoral Researcher with the University of Zurich. He is currently the Head of the Real-Time Data Processing Department, German Aerospace Center (DLR). He is author of various SCI, EI, and Scopus indexed journals and international conferences. His research interests include stereo image processing, deep neural networks, simultaneous localization and modeling (SLAM), and 3D reconstruction.

...